Conference Paper

Hidden Anomalies Detection in Large Arrays of Nuclear Power Plant Operating Data

Chelebiev R. A. and Skomorokhov A. O.

Obninsk Institute for Nuclear Power Engineering of the National Research Nuclear University MEPhI, Studgorodok 1, Obninsk, Kaluga region, 249040, Russia

Abstract

In this paper, it is investigated nuclear power plant operating data which was obtained from reactor main coolant pumps (MCP) of the third isolated generating plant of Kalinin NPP. It is necessary permanent monitoring for state of all pump components since breakdown of a reactor coolant pump leads to substantial economic losses. It is installed over 50 sensors of different control systems at the every MCP. Received data is stored but it is not analysed for the purpose of discovering joint dependencies between equipment pieces and unobvious, hidden trends of accident propagation. In this work, it was proposed techniques for detection of hidden anomalies and MCP operating regularity based on factor analysis, clustering and linear regression models. It was written a Python script which automates necessary calculations.

Keywords: principal component analysis, factor analysis, sammon mapping, robust statistical methods, main coolant pump (MCP).

Corresponding Author:
Chelebiev R. A.
rkchelebiev@gmail.com

Received: 23 December 2017
Accepted: 15 January 2018
Published: 21 February 2018

Publishing services provided by Knowledge E

Selection and Peer-review under the responsibility of the AtomFuture Conference Committee.

1. Introduction

MCP is a composite pump aggregate that is consist of an impeller, an arbor, a container, ball bearings, outer sealings, an electric engine, a heat-exchange unit and so on. It is necessary to permanently monitoring for state of all pump components to find deviations from a normal operation at the very early stage to prevent possible breakdown of a reactor main coolant pump and guarantee safe conditions for reactor active core. It is installed over 50 sensors of different control systems at the every MCP. Received data is stored but it is not analysed for the purpose of discovering joint dependencies between equipment pieces and unobvious, hidden trends of accident propagation.

2. Input data

Analysed operating data arrays are described in the work [1] and includes 3556 observations of 19 temperature sensors for each MCP. Source data is prepared as four

m×*n* matrices of filtered and averaged over one hour values of detectors readings for each MCP (*m* – observations number, *n* – number of the temperature detectors). Consequently each line of the matrix is MCP condition for a given time moment, which is presented by *n*-dimensional vector.

Operational modes have distinctive features as follows [2]. The four MCP operate in a parallel way and in similar conditions and have close operating parameters such as revolutions per minute, temperatures of different minor details, vibration levels.

Time change of each MCP condition can be determined by a common cause associated with the behavior of the reactor facility in the aggregate or individual processes and deviationы for each MCP.

Reactor coolant pump operation abuse is a rare event. Probability of simultaneous breakdown of two or more MCP is substantially lower than occurrence probability of one anomalous MCP.

Therefore, it can be assumed that if all four working MCP behave in a similar way than we are dealing with a normal operational mode. If one of four MCP behavior (that operate in a parallel way) is different from the other the MCP then its condition is anomalous.

3. Mutual dependence structure of the detector readings

3.1. Sammon mapping

Sammon mapping is a nonlinear method to project initial multidimensional data on a plane to allow visual analysis of data points distribution, clustering and finding outliers. Sammon mapping can be formulated as follows. Let $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N$ denote *N* points in *n*-dimensional space and $\vec{y}_1, \vec{y}_2, \ldots, \vec{y}_N$ denote projections onto a plane. Then vector norm of difference between vectors x-*i* and x-*j* (distance) is given as:

$$d_{ij} = \left\|\vec{x}_i - \vec{x}_j\right\| = \left[\sum_{k=1}^{n} (x_{ik} - x_{jk})^2\right]^{\frac{1}{2}} \quad (1)$$

The distance between projected onto a plane points:

$$d^*_{ij} = \left\|\vec{y}_i - \vec{y}_j\right\| = \left[\sum_{k=1}^{2} (y_{ik} - y_{jk})^2\right]^{\frac{1}{2}} \quad (2)$$

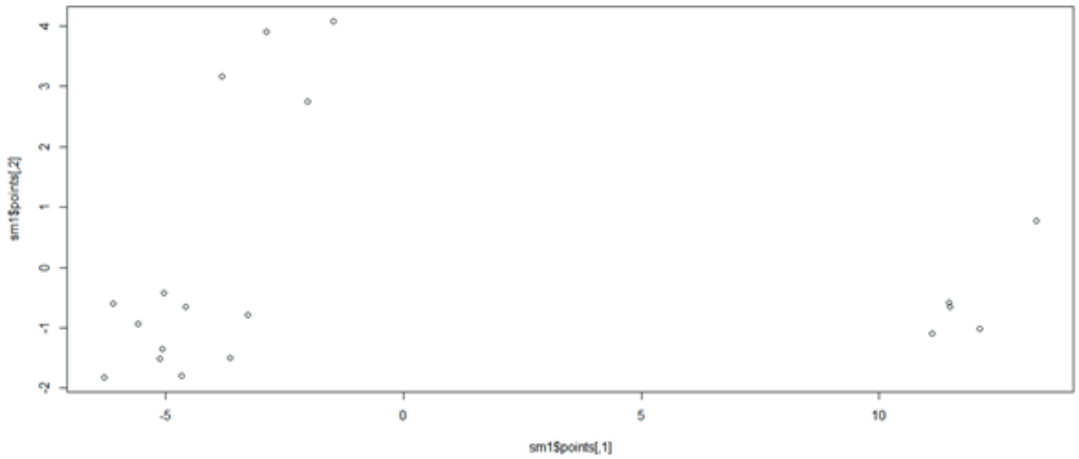

Figure 1: Sammon mapping applied to the transpose data matrix of MCP-1.

Sammon mapping is to minimize a criterion of maximum distance correspondence that is:

$$\varepsilon = \left(\sum_{i<j}^{N} d_{ij} \right)^{-1} \sum_{i<j}^{N} \frac{\left(d_{ij} - d_{ij}^{*} \right)^{2}}{d_{ij}} \tag{3}$$

Described dimensionality reduction method was applied to the transpose data matrix i.e. each sensor is a point in 3556-dimensional space. It was obtained feature grouping for every MCP as illustrated in the Figure 1.

It is appeared a cluster structure of the features. We obtained three clusters using k-means clustering method (see fig. 2). The method is to recalculate each cluster center (it is chosen in a random manner on the first iteration) which is an arithmetic mean of all points relegated to the cluster through proximity measure. In the work was used the function *k-means* implemented in a programming language R for statistical computing [3].

Each cluster corresponds to the same set of detector for every MCP. The first cluster corresponds to sensors with numbers 1-6, 16-19, the second cluster corresponds to sensors with numbers 7, 12-15 and the third one corresponds to sensors with numbers 8-11.

3.2. Factor analysis

Conditional MCP separation of a detector number gave occasion to assume that the features of interest depend of hidden factors caused the observations. To verify this assumption it was calculated eigenvalues of the correlation matrices (as it is done making Principal Component Analysis). Eigenvalues computation was made by eigen

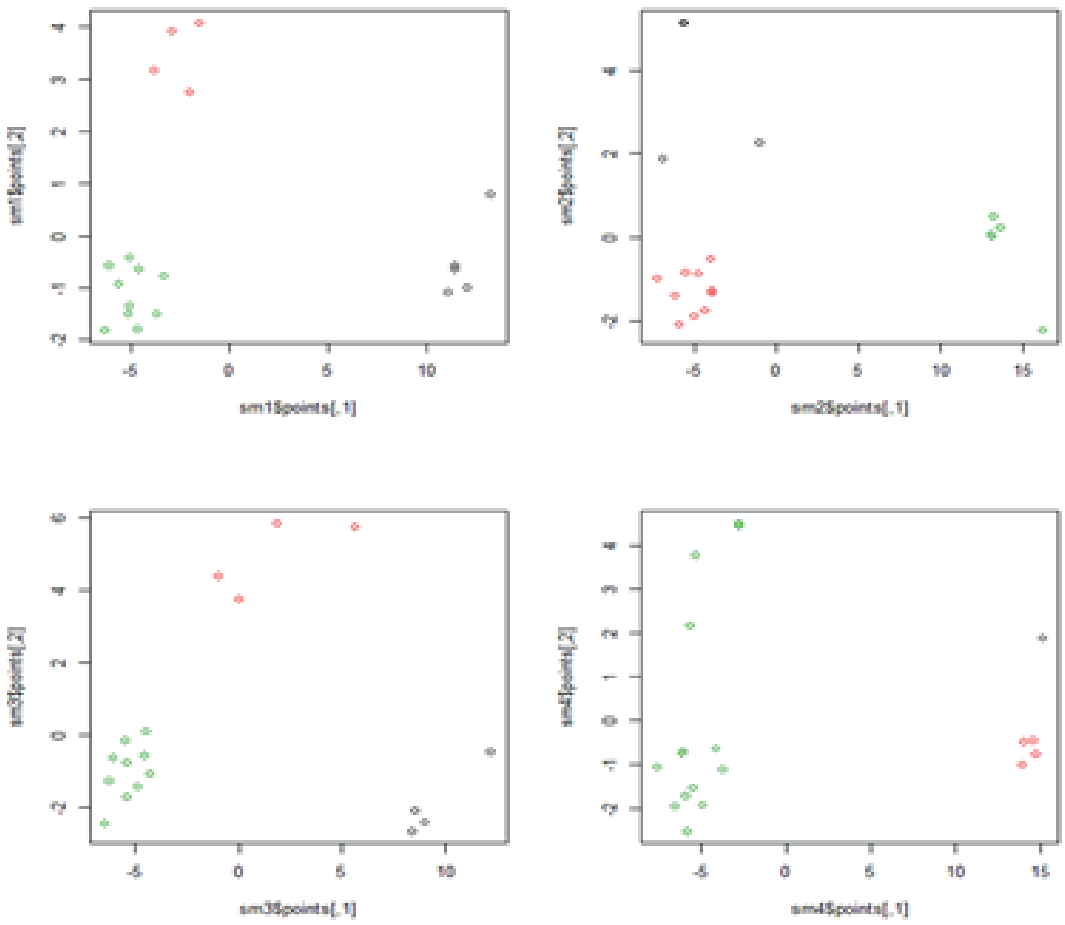

Figure 2: A cluster structure of the temperature sensors readings for all MCP.

decomposition implemented in a programming language R for statistical computing through the function *eigen* [3].

To estimate a number of the principal components it was used Kaiser criterion which is follows: drop all components with eigenvalues under 1. Eigenvalues of each MCP is shown in the Figure 3.

Because only two eigenvalues exceed 1 it was made hypothesis about existence of two hidden factors caused the observation. Hypothesis test was performed by factor analysis. Factor analysis model can be represented as follows:

$$t_i = \lambda_{i1} f_1 + \lambda_{i2} f_2 + \varepsilon_i \tag{4}$$

where $\lambda_{i1,2}$ are loading values of factor 1 and 2 respectively; ε_i are some errors. Obtained result is depicted in the Figure 4.

Loading values which are attributable to sensors 1-6, 8-11, 16-19 of factor 1 noticeably exceed loadings values of sensors 7 and 12-15. Mentioned feature is characteristic of all MCP.

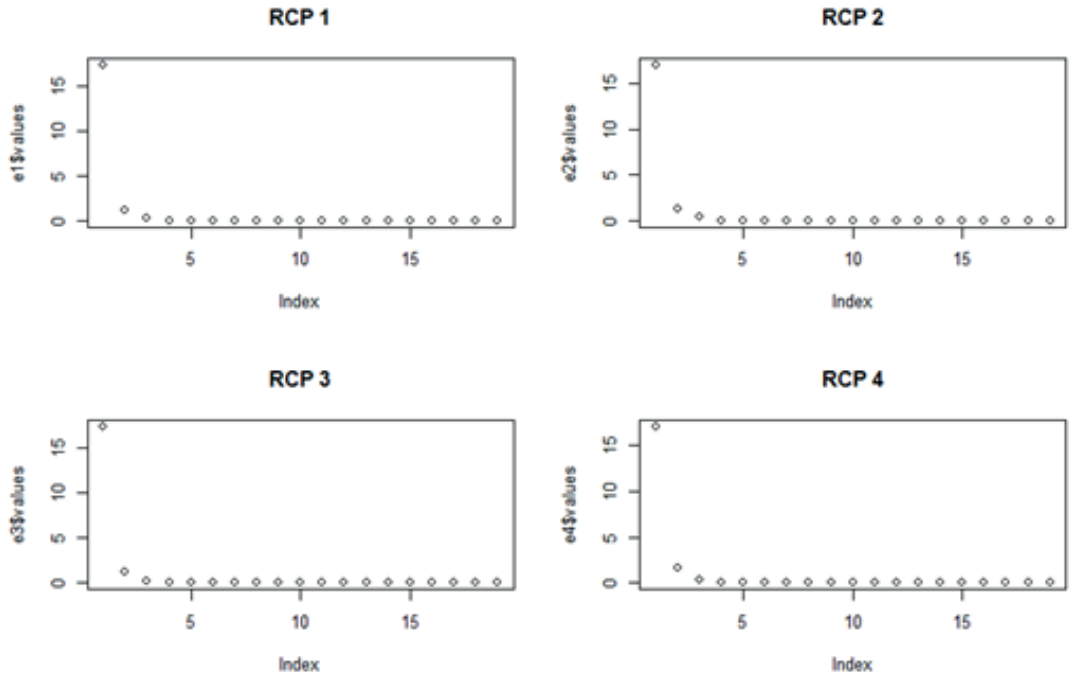

Figure 3: Eigenvalues of the correlation matrices corresponding to each MCP.

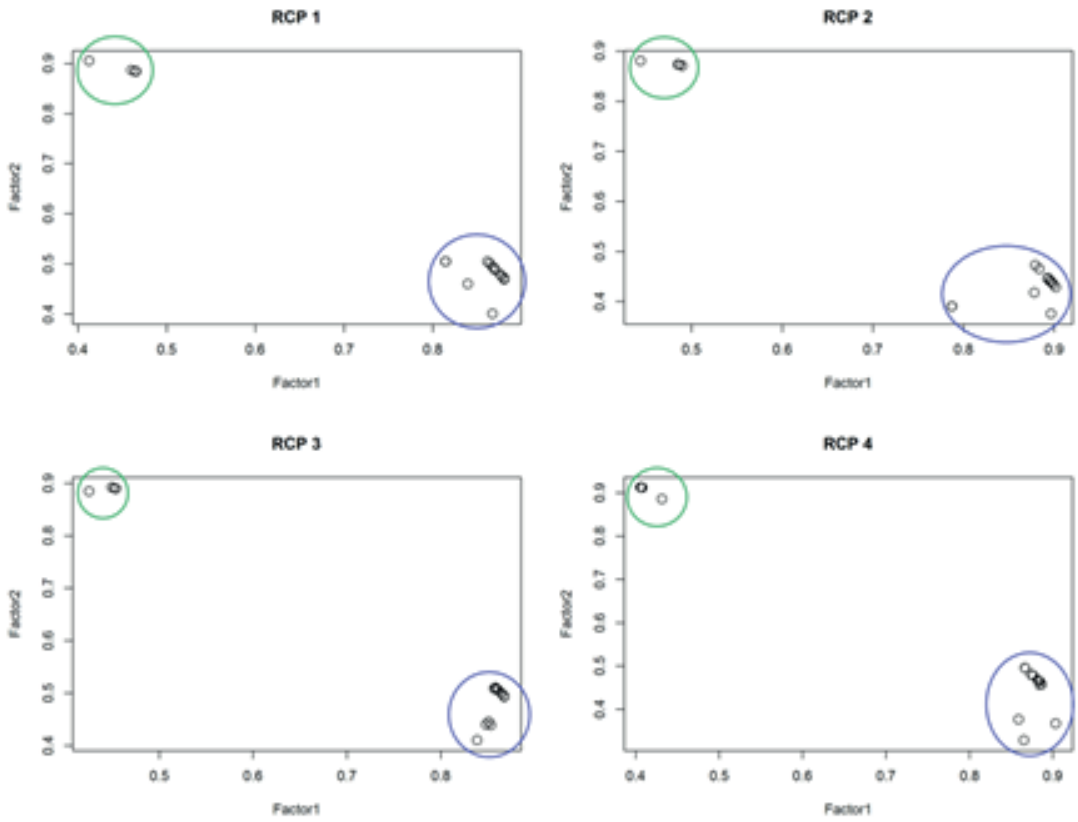

Figure 4: All observations are presented by weighted sum of two hidden factors.

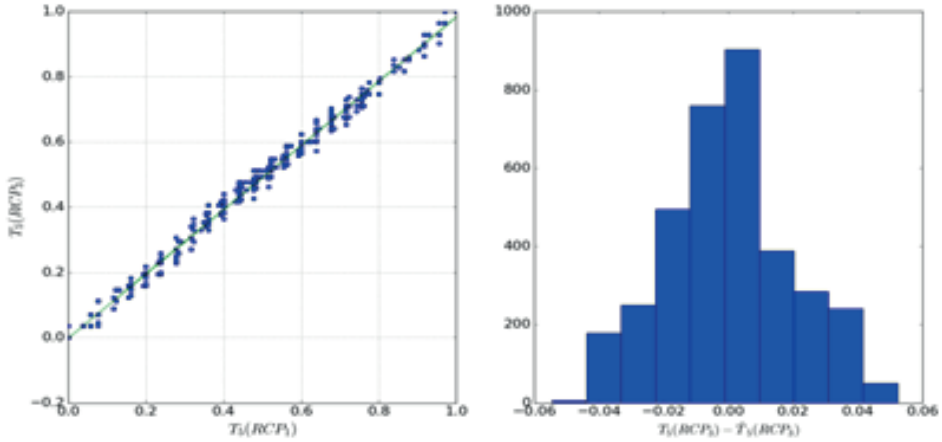

Figure 5: The fifth temperature sensor relationship between MCP-3 and MCP-1.

4. Mutual dependence of the same detectors of different MCH

4.1. Simple linear regression model

The first step of the operating data research is to build linear regression models of detector observations which is belonged to one kind of detector between all possible MCP pairs. The choice of linear regression model is substantiated by the statement above about parallel operating in identical conditions of all MCP. The Figure 5 shows expected dependence of the temperature readings of the sensor 5 between MCP-3 and MCP-1 and histogram of residuals which are a difference between the observations and approximating straight line.

4.2. Outliers presence. Robust regression

Since linear dependence of some detector observations is fractured by outliers which are located sequentially in time, it was chosen robust linear regression model. Usage of least squares criterion leads to incorrect linear model estimates which is sensitive to outliers. For this reason linear model parameters were estimated by the RANdom SAmple Consensus method (RANSAC) that is implemented in a Python library scikit-learn [4]. RANSAC allows to estimate linear regression model parameters without accounting points which are different from main part of data sample (i.e. outliers). The essence of the method is as follows:

- in a random manner it is selected a subset from the original data sample;
- it is calculated linear regression coefficients at the points selected subset;

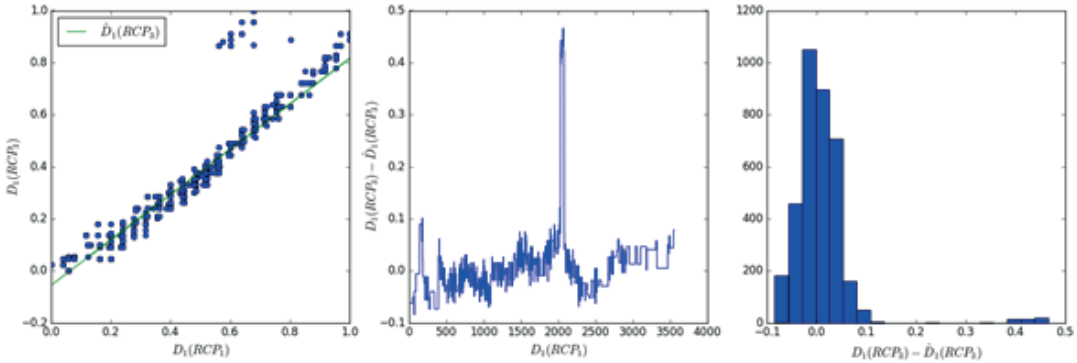

Figure 6: The first temperature sensor relationship between MCP-2 and MCP-1.

- it is stored points number that located in adjacency of the estimated linear model;
- described above steps repeat fixed number;
- the regression model that includes most of the points is accepted as the best one.

For more detailed information on RANSAC method see the work created by Martin A. Fischler and Robert C. Bolles [5].

During the analysis of the all 114 pair relationships the linear behavior is not kept between even numbers and odd MCP numbers. The Figure 6 shows a condition relationship between MCP-2 and MCP-1 of the first temperature sensor. The Figure 7 illustrates common relationships that allow to make a conclusion about a group behavior of MCP.

Outliers detection in automatic mode is based on the work [6] and consists in transforming time measurement series by the formula:

$$\frac{\vec{t} - med(\vec{t})}{1.483 \times med\left|\vec{t} - med(\vec{t})\right|} \tag{5}$$

The threshold value was established on the level equal to 5 by empirical technique. It allowed detect all events of outlier occurrence.

4.3. Linear dependence stratification

In some cases, typical for MCP-3 and 1 the normal distribution of the residuals is broken which consist in splitting of the normal distribution into several disjoint groups. The evaluation of the linear regression parameters corresponding to each of the described groups led to the clarification of straight line stratification which is shown in the Figure

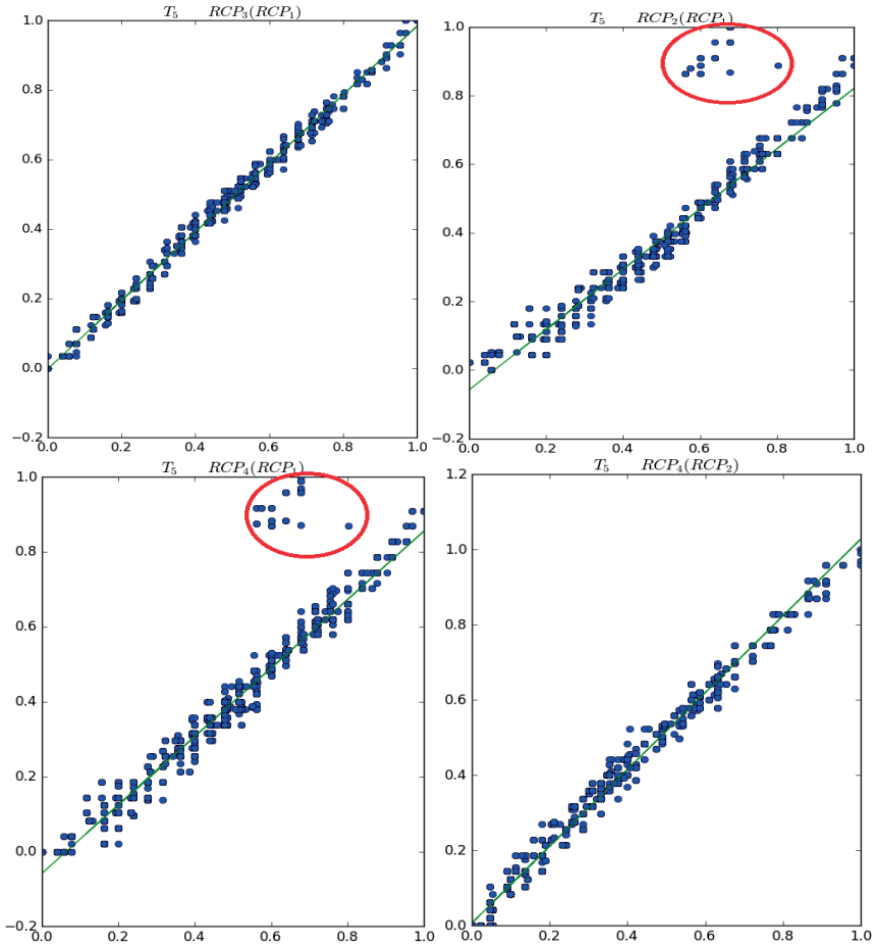

Figure 7: Group behavior of reactor main coolant pumps.

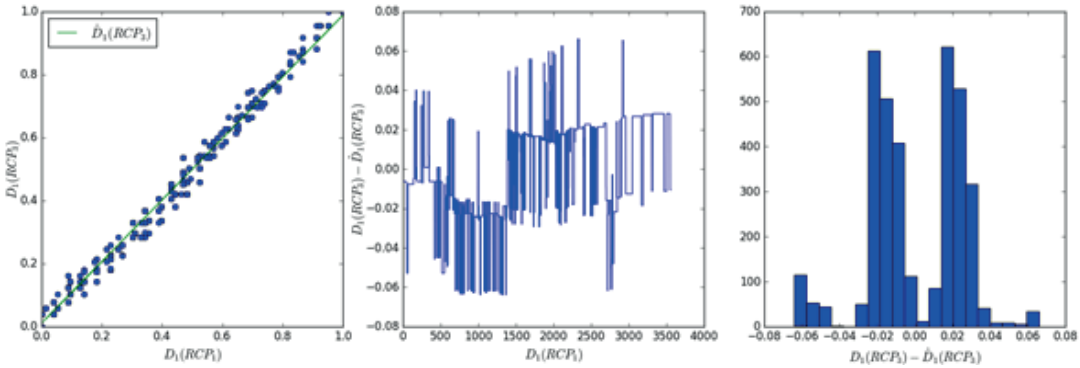

Figure 8: Straight line stratification corresponding to the sensor 1 for MCP-3 and MCP-1.

8. The stratification is manifested in the slopes equality for each point groups and in difference of the free terms values.

Thereby during the analysis it was found the MCP operation has pair behavior: MCP-1 and MCP-3 are not differ from each other whereas operation of MCP-2 and MCP-4 differ from the first group. This feature is typical for most sensors except 6, 7, 10-15.

5. Automation of complex data analysis

Because of a large number of calculations namely the models was built for all possible MCP pair combination (total number is 6) over all 19 detectors (total number of caces is $6 \times 19 = 114$) it was written a script in high-level programming language Python. The presence of a large number of optimized libraries in combination with advantages of Python as a multipurpose programming language makes him a good choice for data wrangling and analysis. Moreover the programming language stands out by a large and active scientific community. This script is to be started from command line in terminal window and in directory, where all data files are located. It runs and automatically create a markdown document with formatted results of all calculations and markdown references to created image files with results of plotting all necessary graphs.

Written program realizes almost all performed analysis. The output of the program is a set of image files and a text report which structures given image files into a logically complete document.

The resulting document may, with the help of pandoc [7, 8], may be converted into a bunch of other formats such as pdf, html, etc.

6. Conclusion

As a result of nuclear power plant operating data research it was recognized pair MCP operating namely MCP operation has a group behavior: MCP-1 and MCP-3 are not differ from each other whereas operation of MCP-2 and MCP-4 differ from the first group.

Also it is found out straight line stratification corresponding to some detectors of MCP-3 and MCP-1. The stratification is manifested in a piecewise linear approximation of mentioned some sensors between MCP-1 and MCP-3. Such stratification allow make an assumption about non registered influence.

The need for a large amount of computations and the construction of the required graphs in the automatic mode caused the creation of a program that implements the above-described analysis of nuclear power plant operating data.

References

[1] Подготовка данных для проведения диагностики состояния ГЦН 3-го блока Калининской АЭС. М.Р. Лапшин, С.Т. Лескин, А.О. Скоморохов. ИАТЭ НИЯУ МИФИ, г. Обнинск.

[2] ОКБ «ГИДРОПРЕСС», 7-я международная научно-техническая конференция «Обеспечение безопасности АЭС с ВВЭР» ДИАГНОСТИКА ГЦН ВВЭР-1000 ПО ДАННЫМ ОПЕРАТИВНО-ТЕХНОЛОГИЧЕСКОГО КОНТРОЛЯ. С.Т. Лескин, В.И. Слободчук, А.С. Шелегов, М.Р. Лапшин, Обнинский Институт Атомной Энергетики, ИАТЭ НИЯУ МИФИ.

[3] R Core Team. 2015. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org.

[4] Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. 2011. "Scikit-Learn: Machine Learning in Python." Journal of Machine Learning Research 12: 2825–30.

[5] Fischler, Martin A, and Robert C Bolles. 1981. "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography." Communications of the ACM 24 (6). ACM: 381–95.

[6] Rousseeuw, Peter J, and Mia Hubert. 2011. "Robust Statistics for Outlier Detection."Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1 (1). Wiley Online Library: 73–79.

[7] Krewinkel, A., & Winkler, R. (2016). *Formatting Open Science: agile creation of multiple document types by writing academic manuscripts in pandoc markdown* (No. e2648v1). PeerJ Preprints.

[8] MacFarlane, J. (2013). Pandoc: a universal document converter. *URL: http://pandoc.org*.